HYDROLOGICAL PROCESSES
Hydrol. Process. **25**, 511–522 (2011)
Published online 23 November 2010 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/hyp.7841

Hydrological field data from a modeller's perspective: Part 1. Diagnostic tests for model structure

Hilary K. McMillan,[1]* Martyn P. Clark,[1] William B. Bowden,[2] Maurice Duncan[1] and Ross A. Woods[1]

[1] *National Institute for Water and Atmospheric Research (NIWA), Christchurch, New Zealand*
[2] *School of Environment and Natural Resources, University of Vermont, Burlington, VT 05405, USA*

Abstract:

Hydrological scientists develop perceptual models of the catchments they study, using field measurements and observations to build an understanding of the dominant processes controlling the hydrological response. However, conceptual and numerical models used to simulate catchment behaviour often fail to take advantage of this knowledge. It is common instead to use a pre-defined model structure which can only be fitted to the catchment via parameter calibration. In this article, we suggest an alternative approach where different sources of field data are used to build a synthesis of dominant hydrological processes and hence provide recommendations for representing those processes in a time-stepping simulation model. Using analysis of precipitation, flow and soil moisture data, recommendations are made for a comprehensive set of modelling decisions, including Evapotranspiration (ET) parameterization, vertical drainage threshold and behaviour, depth and water holding capacity of the active soil zone, unsaturated and saturated zone model architecture and deep groundwater flow behaviour. The second article in this two-part series implements those recommendations and tests the capability of different model sub-components to represent the observed hydrological processes.

KEY WORDS hydrology; rainfall-runoff processes; model structure; data; diagnostic

Received 19 April 2010; Accepted 15 July 2010

INTRODUCTION

The value of multiple field data sets in allowing an observer to develop a perceptual model of a catchment has long been recognized in hydrology. The perceptual model may evolve in response to new data which challenges the current paradigm, leading to a changing collective understanding of the catchment response (e.g. Kirby *et al*., 1991 at Plynlimon; McGlynn *et al*., 2002 at Maimai; Peters *et al*., 2003 at Panola). The development of corresponding conceptual catchment models has been much slower, due to the inherent difficulties of simplifying complex catchment knowledge into a parsimonious model structure (Dunn *et al*., 2008; Soulsby *et al*., 2008; Tetzlaff *et al*., 2008). Many authors have successfully developed individual models of elements of the rainfall–runoff process in well-studied catchments. For example, Birkel *et al*. (2010) simulate saturated area dynamics for the Girnock catchment in the Cairngorns, Scotland; Sidle *et al*. (2001) simulate macropore flow in the Hitachi Ohta Experimental Watershed, Japan; Jencso *et al*. (2009) simulate hillslope-riparian water table connectivity in the Tenderfoot Creek Experimental Forest; Quinn (2004) develops nitrate loss models for the River Ouse and Lehmann *et al*. (2007) model rainfall–outflow thresholds at Panola using percolation theory.

Despite these examples, such models have rarely been combined to build process-based models of a comprehensive range of catchment processes (for example, see Fenicia *et al*., 2008a), nor have these models generally been considered applicable beyond the original experimental location. This difficulty was highlighted by Montanari and Uhlenbrook (2004) and has been described by Beven (2000, 2002a) as the 'uniqueness of place'. It remains a major challenge for the PUB (Prediction in Ungauged Basins) community, who seek to design hydrological models for catchments prior to the availability of extensive field data (McDonnell *et al*., 2005, 2007).

A promising new development to address the challenge of providing tailored, process-orientated conceptual models on a wider scale is through the use of flexible model structures. This approach provides the flexibility to trial different model structures or components, which can be of similar (Clark *et al*., 2008) or varying (Fenicia *et al*., 2006, 2008b) complexity. A flexible model framework might not include every component needed to address a particular perceptual model; however, new components can be added more easily if the software is flexible. The key advantage of a flexible model is that it could adequately represent the hydrologist's perceptual model, using available components, at relatively small cost compared to developing a focused, place-based model. This type of approach might allow more effective interaction between experimentalists and modellers, encouraging the

* Correspondence to: Hilary K. McMillan, National Institute for Water and Atmospheric Research, 10 Kyle Street, Riccarton, Christchurch, New Zealand. E-mail: h.mcmillan@niwa.co.nz

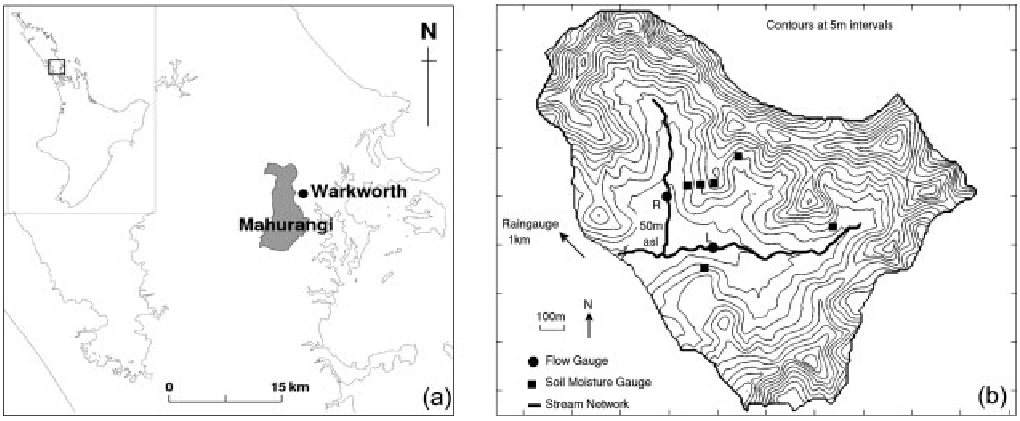

Figure 1. (a) Location map for Mahurangi catchment in North Island of New Zealand and (b) detailed map of Satellite catchment, lying at the eastern point of Mahurangi catchment, showing flow gauges at Satellite Right and Left catchments (marked R and L, respectively) and soil moisture measurement sites

use of field data to infer the choice of model structure as a routine part of hydrological modelling applications.

The objective of this two-part study is to provide guidance on the interpretation of common types of field data for selection of appropriate hydrological model structures. It significantly develops previous work on perceptual and conceptual model building in two ways: (i) It demonstrates how different sources of field data can be used to build a synthesis of dominant hydrological processes and provides recommendations for representing those processes in a time-stepping simulation model and (ii) it implements those recommendations and comprehensively tests the capability of different model sub-components to represent observed hydrological processes. We suggest that an integrated strategy of analysing field data, and analysing model behaviour with respect to process representation, is essential to enable hydrologists to interpret the effects of using different model structures.

In this article, we use precipitation, soil moisture and flow data, both in turn and in combination, to test hypotheses and draw conclusions about process representation within a hydrological model structure. The analysis is based around a case study catchment in New Zealand; however, our aim is to develop methods that are widely applicable. Different analyses or signatures of the same data source can be used to target different components of model structure. The ultimate aim is to use as many data sources as possible to build a complete conceptual model of the target catchment. These structural 'diagnostic tests' draw inspiration from the idea of diagnostic signatures for model evaluation introduced by Gupta *et al*. (2008), who suggest the use of multiple theory-based performance measures to allow identification of relevant model components or parameters. We undertake the field data analyses in conjunction with model selection and sensitivity testing using the Framework for Understanding Structural Errors (FUSE) (Clark *et al*., 2008), to provide guidance on possible model formulations and ensure that model recommendations are linked to accepted lumped catchment model components for easy application and transferability. The companion article uses a variety of FUSE models to test each modelling recommendation and compare analyses of modelled and measured responses using the same range of diagnostics.

The article is structured as follows: the Section on Hydrological Research at Mahurangi and Satellite Catchments describes the catchment and field data available, with information on initial perceptual models of the catchment, the Section on Diagnostics for Conceptual Model Structure shows analysis of the field data (split by flow, soil moisture and water balance analyses), followed by interpretation of the data in terms of catchment process and implications for model design, and the Section on Discussion considers aspects of the results, including scaling issues and freedom in model structure choice.

HYDROLOGICAL RESEARCH AT MAHURANGI AND SATELLITE CATCHMENTS

Overview

Mahurangi catchment is located in the North Island of New Zealand (Figure 1a). The climate is generally warm and humid, with mean annual rainfall of 1628 mm and mean annual pan evaporation of 1315 mm. The Mahurangi River Variability Experiment (Woods *et al*., 2001) ran from 1997 to 2001 and investigated the space-time variability of the catchment water balance. Data from 29 nested stream gauges and 13 rain gauges were complemented by measurements of soil moisture, evaporation and tracer experiments. Within the Mahurangi catchment, several intensive field campaigns have been conducted in Satellite catchment, a sub-catchment of the Mahurangi (Figure 1b). Data from Satellite catchment are used in all the analyses that follow.

Satellite catchment is part of a dairy farm, comprising predominantly pasture with some small areas of scrub, on gently undulating terrain. Elevations range from 50 to 115 m above sea level. Approximately 80% of the catchment forms hillslopes with silty clay loam soil. The remaining 20% forms lowland valleys with alluvial fill soil of a relatively deep profile and high clay content. Both soil types are subject to cracking during dry periods.

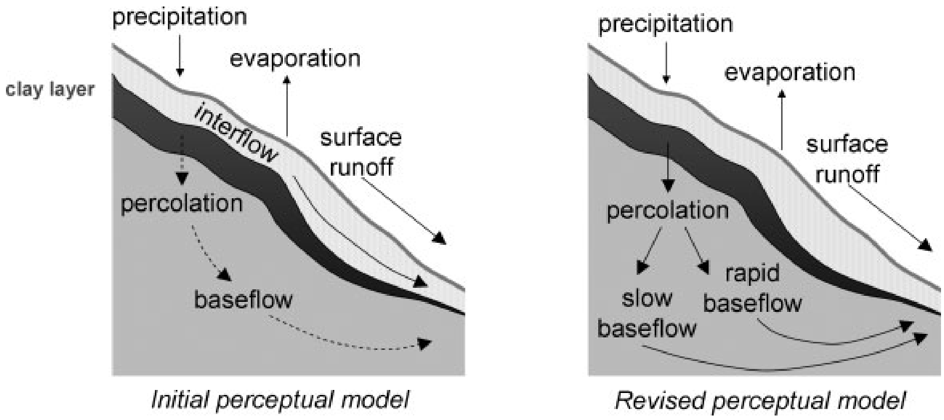

Figure 2. Evolving perceptual models of hillslope processes at Satellite catchment

The catchment is drained by two streams, splitting it into Satellite Right (0·251 km^2) and Satellite Left (0·573 km^2).

Data availability

Both Satellite Right and Left streams were gauged with v-notch weirs; data were recorded at 2-min intervals for three water-years 1998–2001. Streamflows were estimated using weir formulae, checked against current meter readings. Tipping bucket rainfall measurements are available 1 km northwest of Satellite catchment.

Soil moisture was measured at six locations in Satellite catchment, including three aligned on a hillslope transect in Satellite Right (Wilson *et al*., 2003; Western *et al*., 2004). Measurements were made at 30-min intervals for 34 months, at two soil depths: the first at 0–300 mm and the second over the 200 mm of soil at the bottom of the soil profile—the deeper vertical measurement of soil moisture was made at 300–500 mm at the lower hillslope site, 320–520 mm at the middle hillslope site and 600–800 mm at the upper hillslope site. The sensors used were Campbell Scientific CS615 Time Domain Reflectometry probes.

In addition to the continuous soil moisture series, manual measurements were also taken for depths up to 150 cm. These manual measurements were taken in the same locations as the continuous measurements, using an access tube and neutron probe moisture meter. These measurements were taken at eight depths, at 2- to 6-week intervals from 1998 to 2002.

Perceptual models of Satellite catchment

Previous research at Satellite catchment has resulted in evolving understanding of many different aspects of catchment behaviour and process. Western *et al*. (2004) used geostatistical techniques to examine the distribution of soil moisture at Satellite catchment and found significant variation at small scales, which could not be explained by topography. Instead, soil texture and macroporosity were suggested as controlling factors. It was also noted that correlation lengths do not change seasonally at Satellite catchment as wetting and drying occurs all year round; instead, the authors suggest that deeper lateral flow paths may control flow.

Tracer studies reported by W. B. Bowden (2009, personal communication) challenged the prevailing view of the time of the role of shallow soil moisture as a control on flow response. The work describes an initial perceptual model in which flow paths were confined primarily to the upper 30–50 cm of soil, impeded by an underlying clay layer. To test these hypotheses, a multiple-tracer experiment was performed in which bromide was applied to the top of the hillslope and both chloride and deuterated water were applied to the lower slope. However, tracers were never detected in stream water at the base of the hillslope (over a period of 2 months after application) and often bypassed sampling devices within the soil matrix, presumably via preferential flow paths. The fast response component of runoff for this site was therefore suggested to be due to a combination of direct channel interception and local runoff from the near-stream margin, while the majority of hillslope precipitation percolates downwards to the saturated zone. This hypothesis is consistent with previous work in a small, pasture, NZ catchment, which concluded that quickflow is derived from saturation excess flow rather than sub-surface flow (McColl *et al*., 1985). Figure 2 illustrates the change in perceptual model resulting from the experimental work. Similar findings, in which initial hypotheses of shallow flow are contradicted by tracer or isotope measurements suggesting deeper flow paths, are not unusual and have been reported by other authors (Sklash *et al*., 1976; Bestland *et al*., 2009); and evidence of dominant vertical drainage paths and significant deep groundwater contributions to streamflow has also been found at a variety of NZ locations (Rosen *et al*., 1999; Stewart *et al*., 2007; Stewart and Fahey, 2010).

The need for additional, deeper flow pathways to reproduce catchment response has been suggested by previous modelling studies of the wider Mahurangi catchment. Atkinson *et al*. (2003a,b) found that the most crucial addition to a simple catchment model with a single storage reservoir, in order to improve model performance in predicting hourly flow values, was a hillslope representation using a distribution function to imitate the effect of

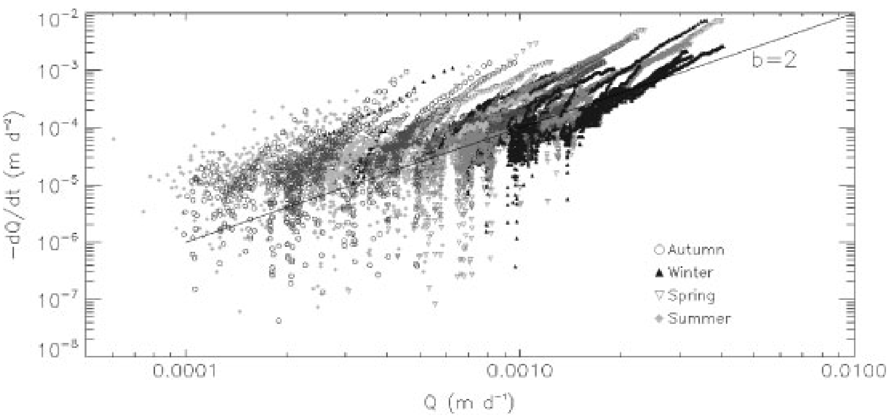

Figure 3. Recession relationships between flow (Q) and flow time derivative (dQ/dt) for Satellite Right catchment, by season

including multiple parallel storage reservoirs. A lumped model including this feature achieved accuracies close to those of a distributed model, where spatially distributed rainfall was used to drive a model with spatially variable parameters. The exception was during dry summer conditions when the distributed model showed an improved ability to capture the catchment dynamics, indicating more complex behaviour when the first part of the precipitation volume is used in 'wetting up' the catchment. Chirico *et al.* (2003) also fitted a fully distributed model to the Mahurangi catchment and similarly found that it was necessary to increase the complexity of the original power-law formulation used to describe the depth-averaged soil lateral transmissivity in order to fit the observed nonlinear flow recession behaviour. The new formulation used the sum of two power-law components, effectively adding an additional flow pathway to the model to allow two parallel, nonlinear storage–discharge contributions with different characteristic response times.

DIAGNOSTICS FOR CONCEPTUAL MODEL STRUCTURE

In this section, we describe a series of diagnostics based on different aspects of the field data collected in Satellite catchment. Each diagnostic targets a particular data source or combination of sources and is interpreted in terms of catchment process understanding and description. The implications for model structure or parameterization then follow. The choice of diagnostics is guided by an evaluation of the components of a typical hydrological model, which are related to the data source(s) in question, the decisions required in selecting a structure for those components and the ability of field data interpretations to guide those decisions.

Diagnostics based on flow data

An established method to study the storage–discharge behaviour of a catchment is via recession analysis (Tallaksen, 1995; Lamb and Beven, 1997; Wittenberg, 1999; Kirchner, 2009). This technique examines the relationship between discharge and its time derivative:

$$-dQ/dt = f(Q) \tag{1}$$

In a conceptual hydrological model, this relationship is uniquely defined by the number, structure and initial conditions of lower zone reservoirs. Therefore, once a model has been selected, the resulting form of the relationship can be compared against measured data (e.g. Clark *et al.*, 2009). McMillan *et al.* (2009) reported the results of recession analysis carried out in the Satellite Right catchment, using the accumulated volume method of Rupp and Selker (2006) to remove noise at low flows. The analysis is illustrated in Figure 3 and shows several key features. Firstly, there is no unique $Q - dQ/dt$ relationship; the relationship varies according to season. Therefore, it follows that there is no unique storage–discharge relationship, and hence a single storage reservoir is insufficient to represent catchment behaviour. Instead, multiple reservoirs are required, such that the proportion of flow originating from each reservoir may vary seasonally. This finding accords with the work of Harman *et al.* (2009a), who found that recession characteristics are sensitive to the recharge history of the catchment.

A second diagnostic is based on the degree of nonlinearity of the recessions. Figure 3 shows that both the initial part and the tails of the recessions are highly nonlinear, with the $Q - dQ/dt$ gradient greater than 2. Where the gradient exceeds 2, the behaviour cannot be replicated with a single (finite, positive capacity) nonlinear reservoir with exponential or power-law behaviour (Rupp and Woods, 2008; Clark *et al.*, 2009). Instead, the behaviour may be accounted for by multiple linear reservoirs or by the hydraulic response of a hillslope undergoing combined saturated and unsaturated flow (Harman *et al.*, 2009b; Szilagyi, 2009). In this case, we seek a conceptual model representation in terms of combinations of reservoirs and hence require at least two nonlinear reservoirs or at least three linear reservoirs. These combinations are the minimal requirements to allow multiple reservoirs to be active throughout the recession.

Diagnostics based on soil moisture data

Soil moisture time series. Soil moisture time series are available in the Satellite Right catchment at three locations and at two depths. Figure 4 shows these data,

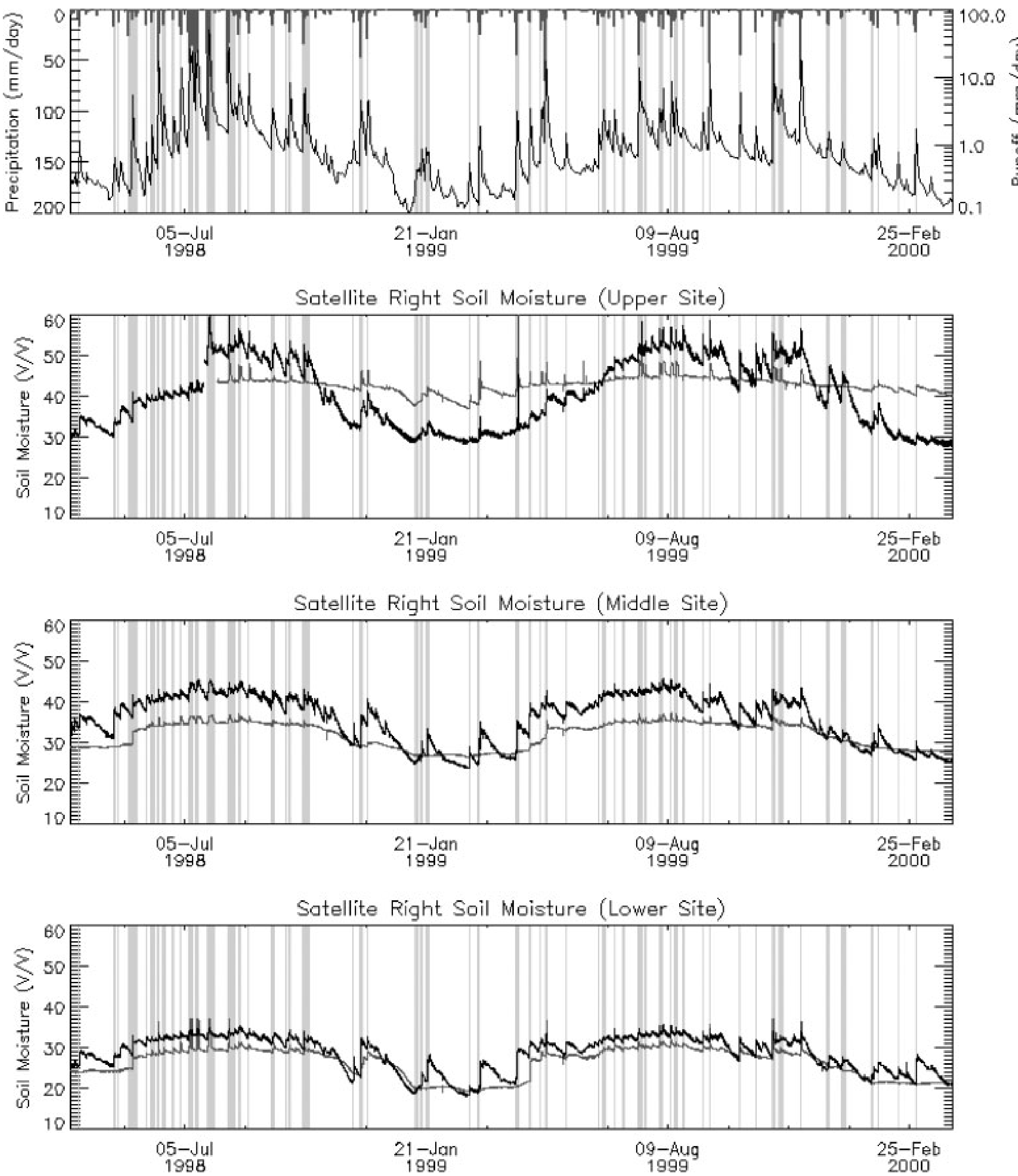

Figure 4. Time series of rain, flow and soil moisture data at three sites in Satellite Right catchment. Upper panel shows precipitation (upper line) and flow on a log scale (lower line). Lower three panels show soil moisture (% v/v) for upper (black line) and lower (grey line) soil layers. Pale vertical lines denote storm periods

together with rainfall and flow, over a 3-year period. Without the requirement for further analysis, these time series can be used to draw simple conclusions regarding the response of the upper soil layers in the catchment.

Examination of the soil moisture series for the upper soil layer in the middle and lower hillslope locations shows that the soil remains above field capacity for only very limited time after rainfall (field capacity, as visually estimated from the time series as the winter 'equilibrium' value for soil moisture, is at 42% v/v for the middle hillslope and 33% v/v for the lower hillslope location). Since the time taken for the upper soil layer to return to field capacity after a rainfall event is too short to invoke ET as the mechanism, rapid horizontal or vertical water movement must occur. The implications from this observation are therefore that the model should allow either a high vertical drainage rate or rapid interflow near to the surface to allow rapid draining of the upper layer.

The soil moisture series show significant differences in behaviour between the upper layer (0–30 cm below

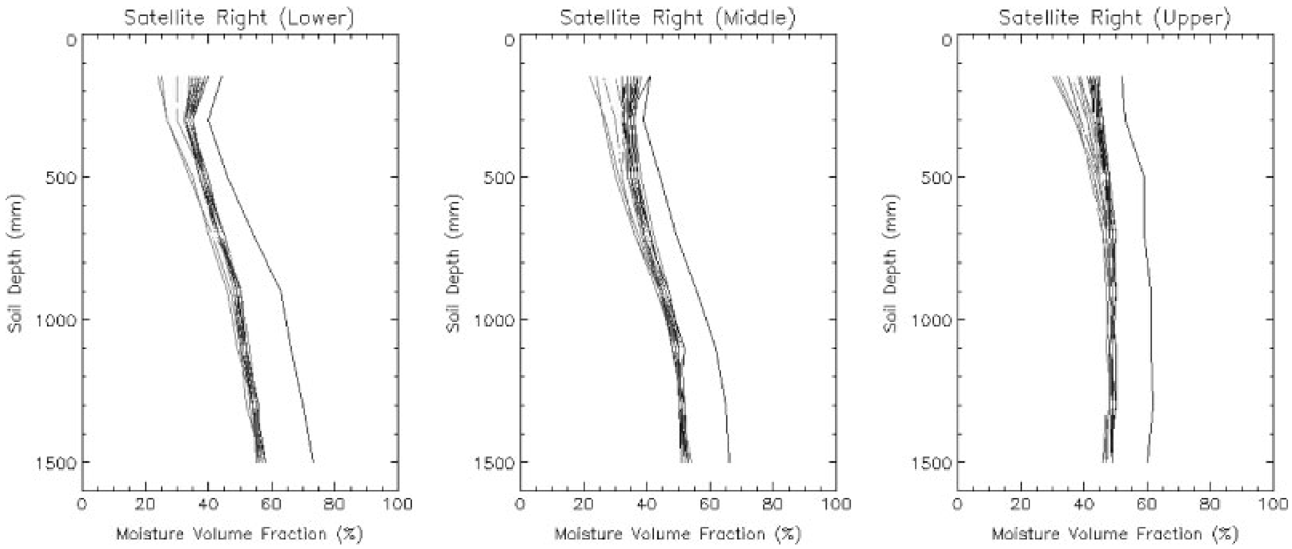

Figure 5. Soil moisture depth profiles constructed from neutron probe measurements for three hillslope sites in Satellite Right catchment. Lines denote individual measurement days

surface) and the lower layer (at the bottom of the soil profile). The lower layer has a delayed wetting-up response at the start of the wetter winter season (for example, during autumn 1998 and autumn 1999 at the middle and lower sites). There is also reduced annual variability of soil moisture in the lower layer, which is typically wetter than the upper layer in summer but drier than the upper layer in winter. These two observations are indicative of a requirement for the conceptual model to include multiple soil layers in the unsaturated zone [e.g. as used in the Precipitation–Runoff Modelling System (Leavesley *et al.*, 1983)]. While variations in behaviour with depth are not sufficient evidence to require additional model complexity *per se* (as the model is necessarily a simplification of the true catchment behaviour), two layers operating independently are likely to be needed to allow the water balance to be maintained through sufficient summer ET from a shallow and hence easily wetted upper layer in the unsaturated zone (Guswa *et al.*, 2004).

The soil moisture series can also be used to learn about ET behaviour and model formulation. During winter months, both upper and lower soil moisture sensors are close to field capacity after a storm event, but only the upper sensor moisture falls between storm events, suggesting that ET demand is satisfied from the upper part of the soil. However, during the summer months when the upper sensor moisture content is significantly lower than field capacity, the lower sensor moisture is also depleted between storm events due to ET (e.g. Figure 4, upper site). We therefore hypothesize that a model of the catchment should use a sequential evaporation scheme, whereby demand is preferentially met by an upper soil layer; then, unsatisfied demand is met from a lower soil layer. This hypothesis will be tested in the companion article.

Soil moisture depth profiles. In addition to the continuous soil moisture series, manual measurements were also taken in the same locations using a neutron probe for depths up to 150 cm. These measurements were taken at intervals between 2 and 6 weeks, from 1998 through 2002, at eight depths (150, 300, 500, 700, 900, 1100, 1300 and 1500 mm from the soil surface). The resulting soil moisture profiles are shown in Figure 5. The profiles show that variability in soil moisture reduces with depth; active variability occurs up to depths of approximately 1 m.

These results can be used to calculate with reasonable accuracy the maximum water content of the soil, which is required as a parameter in many soil models and in turn controls the variability of soil moisture. In Satellite Right catchment, given variability in tension storage of approximately 15% (estimated from the neutron probe measurements, Figure 5), and making a qualitative assessment that tension storage comprises approximately 50% of total storage (Figure 4), the maximum water content should be approximately 300 mm.

None of the soil moisture profiles shows influence from the saturated zone (this would be evidenced by a kink in the profile), suggesting that the water table remains at depths greater than 150 cm. We also deduce that substantial evapotranspiration from the saturated zone is unlikely since plant rooting depths in this pastoral landscape would typically be confined to the upper 50 cm of soil.

Diagnostics based on water balance

The water balance characteristics of the catchment were considered on a per-storm basis. This allowed us to focus attention on the catchment response conditioned on pre-storm wetness conditions and also on the characteristic response timescales of the catchment. To do this, the time series of rainfall and flow were divided into individual storm events. Storm events were objectively identified from aggregated hourly precipitation data as follows:

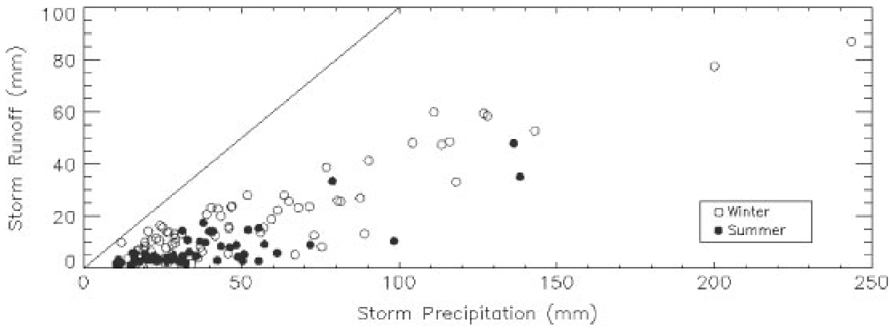

Figure 6. Relationship between storm precipitation depth and storm runoff depth, during winter (open circles) and summer (filled circles). Black line indicates 100% runoff

1. The start of the storm is identified as when both (i) rainfall in a given hour is greater than 0·5 mm/day (parameter *xinter*) and (ii) mean rainfall over the following 24 h (parameter *istorm*) is greater than 10 mm/day (parameter *xstorm*).
2. The end of the storm is identified when the maximum hourly rainfall over the following 36 h (parameter *iinter*) is less than 0·5 mm/day (parameter *xinter*).

The four parameters (*xinter, xstorm, istorm* and *iinter*) were specified based on visual inspection of the data. For calculations of flow depth and flow timing during an event, the event was deemed to end no more than 5 days after the last rainfall (for comparison: typical duration of visible raised channel discharge is of the order of 24 h). Using a fixed length for storms is necessary to evaluate the impact of different model parameters on simulations of runoff volume and runoff timing.

Runoff response to precipitation. The first analysis was simply to compare storm precipitation depth to storm runoff depth. A graph of these values is shown in Figure 6, with storm events additionally identified by season. For summer (October–March) storms, there is a threshold of approximately 30-mm rainfall depth, below which storm runoff is close to zero. In winter (April–September), no threshold exists and runoff is recorded even for the smallest storm events. Threshold responses for precipitation have been reported previously by Tromp-van Meerveld and McDonnell (2006a,b) at Panola catchment and interpreted as a 'fill-and-spill' process by which depressions in the bedrock at the soil–rock interface must be filled before downslope flow (and hence channel runoff) occurs. Tromp-van Meerveld and McDonnell (2005) and Western *et al*. (2004, 2005) discussed alternative theories for the importance of thresholds on controlling runoff, with soil moisture and transient saturation discussed as competing hypotheses for sub-surface lateral flow. Alternative conceptualizations for threshold behaviour have also been proposed such as connection of lateral preferential flow pathways (Sidle *et al*., 2000).

At Satellite catchment, the threshold in runoff response is functionally different from that observed at Panola because it occurs only in summer. We therefore attribute the control to shallow soil moisture (which has a strong seasonal cycle) rather than bedrock topography. As an alternative hypothesis, it is possible that the water table rises above the bedrock depressions in winter and hence no threshold is provided. However, field observations showed that there is no well-defined soil–bedrock interface to produce depressions: The area has never been glaciated and total soil depth is large (∼10 m) with a gradual transition to bedrock (M. Duncan, 2010, personal communication). Instead, we suggest that the shallow soil layers do not transmit water (laterally or vertically) until a threshold moisture content is reached. The modelled soil should therefore have sufficient depth to allow the 30-mm initial losses to be absorbed before vertical drainage begins. It is instructive to compare the 30-mm precipitation threshold for runoff response with the estimated value of 300 mm of active storage derived from the neutron probe data (see Section on Diagnostics Based on Soil Moisture Data). While we cannot be certain of the processes contributing to the order of magnitude difference between the two depths, we hypothesize that it is due to the crucial role of spatial variability of soil moisture on catchment response. Runoff is likely to be activated at the catchment scale at a lower threshold due to contributions from areas of shallow or highly structured soil not captured by the localized soil moisture measurements.

Runoff ratio. The threshold response to storm precipitation according to initial soil moisture is examined from a different perspective in Figure 7, which shows runoff ratio as a function of storm precipitation and soil moisture at the start of the storm. The figure shows that soil moisture has a much greater control on the runoff ratio than storm precipitation: Different storm depths are associated with a range of runoff ratios, but runoff ratio does depend strongly on the initial soil moisture at the start of the storm. We suggest that precipitation depth controls runoff ratio only indirectly via filling of soil moisture stores before runoff commences during dry summer periods.

The absolute value of the runoff ratio provides another diagnostic of the catchment response to rainfall. The ratio is always below 0·6 (the single point at 0·8 is an outlier caused by elevated flow levels remaining from a previous storm); and even for storms of greater than 100 mm

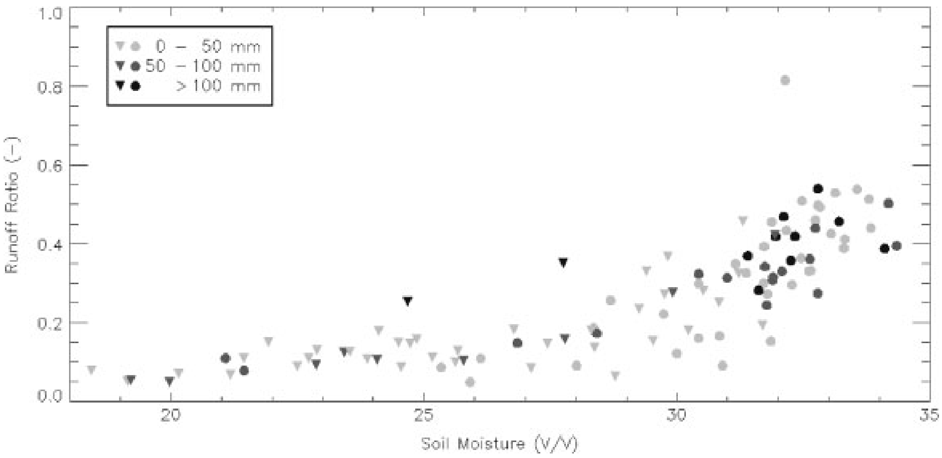

Figure 7. Relationship between initial soil moisture at the start of the storm and the storm runoff ratio, for summer (triangles) and winter (circles). The tone of the symbols denotes total storm precipitation (see legend)

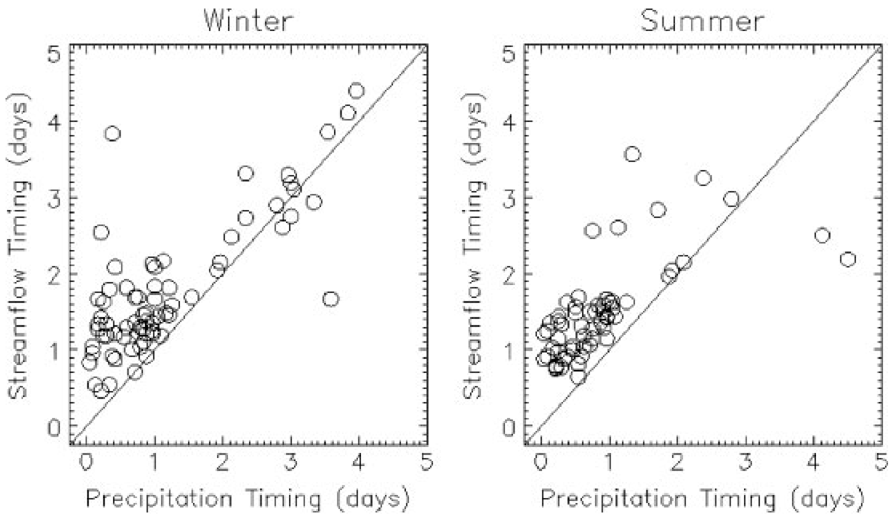

Figure 8. The time since the start of the storm for which 50% of the total storm precipitation and streamflow was observed, for winter (left plot) and summer (right plot)

rainfall, the ratio is often lower than 0·5. Given that the part of the rainfall depth falling into the channel and saturated areas is likely delivered rapidly to the stream, the figures show that the greater part of the rain falling onto the hillslopes does not reach the stream during the storm event. Since any soil moisture above field capacity has been dissipated 5 days after rainfall (Figure 4), this rainfall depth must be either lost to ET (unlikely to be a significant fraction during storm events or in winter), stored as residual soil moisture (likely for cases of low storm rainfall and very low runoff ratio) or percolate to the saturated zone. The catchment model should therefore allow rapid vertical drainage of soil moisture in excess of field capacity to a baseflow reservoir with a dominating slow response component which has a time constant of weeks or longer.

Runoff timing. The lag time between rainfall and runoff centroid is analysed in Figure 8. For each storm event, the time in days between 50% of storm precipitation having fallen and 50% of storm runoff produced is plotted. For a range of lag times (i.e. short intense storms and longer low-intensity storms), the average lag time is around 0·5 days. No significant trend in lag time was found due to precipitation depth. Note that this lag time considers only 50% or less of storm rainfall, which reaches the stream within the storm period.

A lag time of 12 h is relatively slow for a small catchment such as Satellite Right, where the longest flow path lengths are of the order of a few hundred metres. This result suggests that water is moving as slow subsurface flow, where flow path depth and permeability of the regolith or bedrock are more likely to be controlling factors than flow path length alone. We therefore suggest that although the soil profile was found to drain quickly (see Section on Diagnostics Based on Flow Data), this drainage should represent vertical drainage to the saturated zone rather than interflow. It is possible that low lateral conductivity in the upper soil could also cause this effect; however, previous tracer studies (refer to Section on Perceptual Models of Satellite Catchment) also support an absence of significant interflow. This conclusion also suggests that multiple sub-surface pathways are required in the catchment model such that water in the saturated zone may take times ranging from less than 0·5 day to greater than 5 days to reach the channel after a storm event.

DISCUSSION

The use of diagnostics for model structure

We propose in this article a collection of analyses or diagnostics which use commonly collected field data types to guide hydrological model structural choice. For reference, these are summarized in Table I.

Where different data sources were available, for example, on water table depth, saturated area dynamics or from isotope or tracer studies, more diagnostic tests could be applied to target some of the remaining model building decisions. However, our analyses show what is possible when using only time series data for precipitation, flow and soil moisture, which are standard tools for wide-scale catchment-monitoring networks.

Table I. Proposed diagnostics to guide hydrological model structural choice

Data type	Analysis	Model decisions
Flow	Recession analysis • Single/multiple relationship • Degree of nonlinearity	Saturated zone model architecture: number and type of storage reservoirs
Soil moisture	Behaviour above field capacity	Parameterization for drainage of free storage
Soil moisture	Variation in behaviour with depth • ET • Lag of wetting at depth • Strength of annual *vs* storm signal	Unsaturated zone model architecture: number of vertical layers and connectivity of layers ET parameterization
Soil moisture	Temporal variation in depth profiles	Depth and water holding capacity of active soil zone
Precipitation and flow	Threshold in runoff response	Soil water holding capacity Vertical drainage parameterization
Precipitation and flow	Lag between precipitation and runoff centroids	Balance of near-surface and baseflow pathways
Precipitation and flow	Runoff ratio absolute value	Significance and time constant of deep groundwater flow
Precipitation, flow, soil moisture	Control of runoff ratio by precipitation depth and antecedent soil moisture	Threshold behaviour in unsaturated *vs* saturated zone

Table II. Recommendations for Satellite catchment hydrological model

Model component	Recommendation
Unsaturated zone architecture	Multiple cascading soil layers in the unsaturated zone
Unsaturated zone parameter values	Maximum water content of active storage ≈300 mm Threshold storage before drainage occurs ≈30 mm
Evapotranspiration	Sequential ET scheme where demand is met preferentially from the upper soil layer. No ET from the saturated zone
Interflow	Interflow is not a dominant process
Saturated zone architecture	At least two nonlinear reservoirs or three linear reservoirs (or equivalent combination) to allow seasonality and nonlinearity of recession behaviour. Characteristic response time should range from <0·5 to >5 days
Drainage parameterization	Dominant vertical drainage pathway which allows rapid drainage (sub-day) of water when the soil is above field capacity. Drainage occurs below field capacity only as a proxy process for heterogeneity of soils. Drainage not controlled by the saturated zone

An important challenge lies in being able to predict which data presentations or diagnostic methods will be useful for model identification prior to the analysis being carried out. However, this ability is critical if model identification is to become accessible for widespread use. This article starts towards building a toolbox of useful diagnostic tests and we welcome discussion on further diagnostic tests for model structure.

Recommendations for Satellite catchment

Implementation of the collection of diagnostic tests described above allows us to make explicit recommendations for the structure of a hydrological model for the Satellite catchment of the Mahurangi. These model recommendations are tested in the companion article but are summarized in Table II for completeness. It should be noted that calibrated hydrological models using conflicting structures may perform equally well at reproducing flow hydrographs, as measured by typical tests of model performance such as the Nash–Sutcliffe score. However, such models would be expected to perform less well at reproducing the process-based diagnostics suggested here, while retaining physically realistic parameter values, and hence would be less reliable for use in applications concerned with a broader interpretation of catchment behaviour, such as response to land use or climate changes, or contaminant transfer processes.

Scaling

We have suggested diagnostics for both model structure (e.g. number of storage reservoirs) and model parameter values (e.g. maximum soil water capacity). Both these diagnostic types may be affected by problems of scaling, as localized field data are used to draw conclusions about wider catchment response (Blöschl and Sivapalan, 1995; Western *et al*., 2002; Sivapalan *et al*., 2004). Parameter values are particularly susceptible: The approach we used was to look for behaviour that was repeated at different locations in the catchment to indicate consistent function. However, the proposed soil moisture-based quantities are only directly applicable to small scales and their transferal to large-scale lumped conceptual model components should be undertaken with caution. In future, remote sensing techniques that provide integrated spatial information (e.g. Pauwels *et al*., 2009; Minet *et al*., 2010) have the potential to reduce the scale difference between diagnostics and models.

Model structure, despite perhaps representing more fundamental modelling choices, is also affected by issues of scale. For example, threshold behaviour which occurs everywhere in the catchment, but with a varying threshold according to location, may lead to a catchment response in which the threshold is blurred, as was found here with measured *vs* effective field capacity threshold (see Section on Diagnostics Based on Water Balance). Smoothing of threshold behaviour has previously been suggested as providing model equivalence for other threshold-driven but spatially varying processes such as snowmelt, as well as being recommended to remove numerical artefacts (Kavetski *et al*., 2006). Although dominant processes may also differ with scale (e.g. matrix *vs* preferential flow), this is less problematic as our choice of model structures reflects prior understanding of possible process behaviour at the lumped catchment scale.

Measured input and response data, and hence parameter values and diagnostics, are subject to varied sources of uncertainty beyond those originating from scaling issues. These sources may, for example, relate to measurement frequency, equipment calibration and rating curve formulation. Therefore, to fully assess model structure behaviour and reliability against field data, a probabilistic approach to diagnostic testing will be required. Probabilistic diagnostics have previously been suggested as necessary to assess model behaviour against uncertain flow data (McMillan *et al*., 2010) and Thyer *et al*. (2009) present diagnostic measures, such as flow quantile–quantile plots, which directly assess the reliability of the model's predictive limits. However, the development of probabilistic diagnostics which reflect the range and type of physical insights described here remains a challenge for the future.

Model structural choices

By basing our model structural recommendations on commonly used hydrological modelling components [such as those described in FUSE (Clark *et al*., 2008)], we benefit from the accumulated knowledge of previous hydrologists and model builders in terms of successful catchment process representation. The approach of choosing or learning from existing model functionality builds on previous studies which have used a rejectionist framework to assess different model structures against analysis of field data sources (e.g. Vache and McDonnell, 2006), retaining those models which do not contradict observed data. However, there is an associated risk that the range of analyses considered is unconsciously constrained by pre-conceived structural choices. Diagnostics to test for processes which were not included in any of the multi-model structures may not be so easily defined. In this aspect, experimentalists' interpretation of the data is key to more creative thinking in terms of model structure.

The opposite case may also be encountered: Where field data with high temporal or spatial resolution are available, the temptation is to construct a model to mimic the data as closely as possible. However, the hydrological model must always be a simplification of true catchment behaviour, and the modeller's skill lies in understanding where simplifications in model structure can be made without jeopardizing model ability to simulate critical catchment fluxes. In this way, the 'landscape space' can be mapped onto the reduced dimensionality of the 'model space' (Beven, 2002b). The ability to link landscape form to model structure will be essential for the long-term aim of including structural identification within hydrological regionalization algorithms, which are currently hampered by model structural uncertainty (Wagener and Wheater, 2006).

CONCLUSIONS

Current hydrological modelling practice often entails the use of a pre-defined model structure, which is fitted to a specific catchment using inverse modelling for parameter calibration. This 'one-size-fits-all' approach to model structure has been criticized by Savenije (2009) as an engineering concept which is not suitable for the 'art' of hydrological research. This article demonstrates instead how field data (time series of precipitation, soil moisture and flow) can be used to test hypotheses about model structure and so to design a bespoke conceptual model for an individual catchment. Recommendations were made for a comprehensive set of modelling decisions, including ET parameterization, vertical drainage threshold and behaviour, depth and water holding capacity of the active soil zone, unsaturated and saturated zone model architecture and deep groundwater flow behaviour. These suggestions for diagnostic tests for model structure are intended to foster a wider acceptance of the need to both tailor hydrological models for each unique catchment and vary the model structure over larger modelling domains.

REFERENCES

Atkinson SE, Sivapalan M, Viney NR, Woods RA. 2003a. Predicting space-time variability of hourly streamflow and the role of climate seasonality: Mahurangi Catchment, New Zealand. *Hydrological Processes* **17**: 2171–2193.

Atkinson SE, Sivapalan M, Woods RA, Viney NR. 2003b. Dominant physical controls on hourly flow predictions and the role of spatial variability: Mahurangi catchment, New Zealand. *Advances in Water Resources* **26**: 219–235.

Bestland E, Milgate S, Chittleborough D, VanLeeuwen J, Pichler M, Soloninka L. 2009. The significance and lag-time of deep through flow: an example from a small, ephemeral catchment with contrasting soil types in the Adelaide Hills, South Australia. *Hydrology and Earth System Sciences* **13**: 1201–1214.

Beven K. 2002a. Towards a coherent philosophy for modelling the environment. *Proceedings of the Royal Society Mathematical Physical and Engineering Sciences* **458**: 2465–2484.

Beven K. 2002b. Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system. *Hydrological Processes* **16**: 189–206.

Beven KJ. 2000. Uniqueness of place and process representations in hydrological modelling. *Hydrology and Earth System Sciences* **4**: 203–213.

Birkel C, Tetzlaff D, Dunn SM, Soulsby C. 2010. Towards a simple dynamic process conceptualization in rainfall-runoff models using multi-criteria calibration and tracers in temperate, upland catchments. *Hydrological Processes* **24**: 260–275.

Blöschl G, Sivapalan M. 1995. Scale issues in hydrological modeling—a review. *Hydrological Processes* **9**: 251–290.

Chirico GB, Grayson RB, Western AW. 2003. A downward approach to identifying the structure and parameters of a process-based model for a small experimental catchment. *Hydrological Processes* **17**: 2239–2258.

Clark MP, Rupp DE, Woods RA, Tromp-van Meerveld HJ, Peters NE, Freer JE. 2009. Consistency between hydrological models and field observations: linking processes at the hillslope scale to hydrological responses at the watershed scale. *Hydrological Processes* **23**: 311–319.

Clark MP, Slater AG, Rupp DE, Woods RA, Vrugt JA, Gupta HV, Wagener T, Hay LE. 2008. Framework for Understanding Structural Errors (FUSE): a modular framework to diagnose differences between hydrological models. *Water Resources Research* **44**: W00B02.

Dunn SM, Freer J, Weiler M, Kirkby MJ, Seibert J, Quinn PF, Lischeid G, Tetzlaff D, Soulsby C. 2008. Conceptualization in catchment modelling: simply learning? *Hydrological Processes* **22**: 2389–2393.

Fenicia F, Savenije HHG, Matgen P, Pfister L. 2006. Is the groundwater reservoir linear? Learning from data in hydrological modelling. *Hydrology and Earth System Sciences* **10**: 139–150.

Fenicia F, McDonnell JJ, Savenije HHG. 2008a. Learning from model improvement: on the contribution of complementary data to process understanding. *Water Resources Research* **44**: W06419.

Fenicia F, Savenije HHG, Matgen P, Pfister L. 2008b. Understanding catchment behavior through stepwise model concept improvement. *Water Resources Research* **44**: W01402.

Gupta HV, Wagener T, Liu YQ. 2008. Reconciling theory with observations: elements of a diagnostic approach to model evaluation. *Hydrological Processes* **22**: 3802–3813.

Guswa AJ, Celia MA, Rodriguez-Iturbe I. 2004. Effect of vertical resolution on predictions of transpiration in water-limited ecosystems. *Advances in Water Resources* **27**: 467–480.

Harman CJ, Sivapalan M, Kumar P. 2009a. Power law catchment-scale recessions arising from heterogeneous linear small-scale dynamics. *Water Resources Research* **45**: W09404.

Harman CJ, Sivapalan M, Kumar P. 2009b. Reply to comment by J. Szilagyi on "Power law catchment-scale recessions arising from heterogeneous linear small-scale dynamics". *Water Resources Research* **45**: W12602.

Jencso KG, McGlynn BL, Gooseff MN, Wondzell SM, Bencala KE, Marshall LA. 2009. Hydrologic connectivity between landscapes and streams: transferring reach- and plot-scale understanding to the catchment scale. *Water Resources Research* **45**: W04428.

Kavetski D, Kuczera G, Franks SW. 2006. Calibration of conceptual hydrological models revisited: 1. Overcoming numerical artefacts. *Journal of Hydrology* **320**: 173–186.

Kirby C, Newson MD, Gilman K. 1991. *Plynlimon Research: The First Two Decades*. Institute of Hydrology: Wallingford; 188.

Kirchner JW. 2009. Catchments as simple dynamical systems: catchment characterization, rainfall–runoff modeling, and doing hydrology backward. *Water Resources Research* **45**: W02429.

Lamb R, Beven K. 1997. Using interactive recession curve analysis to specify a general catchment storage model. *Hydrology and Earth System Sciences* **1**: 101–113.

Leavesley GH, Lichty RW, Troutman BM, Saindon LG. 1983. Precipitation–Runoff Modelling System: User's Manual. U.S. Geological Survey Water Investigations Report 207.

Lehmann P, Hinz C, McGrath G, Tromp-van Meerveld HJ, McDonnell JJ. 2007. Rainfall threshold for hillslope outflow: an emergent property of flow pathway connectivity. *Hydrology and Earth System Sciences* **11**: 1047–1063.

McColl RHS, McQueen DJ, Gibson AR, Heine JC. 1985. Source areas of storm runoff in a pasture catchment. *Journal of Hydrology (New Zealand)* **24**: 1–19.

McDonnell JJ, McGlynn B, Vache K, Tromp-Van Meerveld I. 2005. A perspective on Hillslope hydrology in the context of PUB. *Predictions in Ungauged Basins: International Perspectives on the State of the Art and Pathways Forward*. IAHS Publication 301; 204–212.

McDonnell JJ, Sivapalan M, Vache K, Dunn S, Grant G, Haggerty R, Hinz C, Hooper R, Kirchner J, Roderick ML, Selker J, Weiler M. 2007. Moving beyond heterogeneity and process complexity: a new vision for watershed hydrology. *Water Resources Research* **43**: W07301.

McGlynn BL, McDonnel JJ, Brammer DD. 2002. A review of the evolving perceptual model of hillslope flowpaths at the Maimai catchments, New Zealand. *Journal of Hydrology* **257**: 1–26.

McMillan HK, Clark MP, Woods RA, Duncan MJ, Srinivasan MS, Western AW, Goodrich D. 2009. Improving perceptual and conceptual hydrological models using data from small basins.. In *Proceedings of the Workshop on Status and Perspectives of Hydrology in Small Basins Held at Goslar-Hahnenklee*, Germany, 30 March–2 April 2009. IAHS Publication 336-08.

McMillan HK, Freer J, Pappenberger F, Krueger T, Clark MP. 2010. Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions. *Hydrological Processes*. DOI: 10.1002/hyp.7587.

Minet J, Lambot S, Slob EC, Vanclooster M. 2010. Soil surface water content estimation by full-waveform GPR signal inversion in the presence of thin layers. *IEEE Transactions on Geoscience and Remote Sensing* **48**: 1138–1150.

Montanari A, Uhlenbrook S. 2004. Catchment modelling: towards an improved representation of the hydrological processes in real-world model applications. *Journal of Hydrology* **291**: p. 159.

Pauwels VRN, Balenzano A, Satalino G, Skriver H, Verhoest NEC, Mattia F. 2009. Optimization of soil hydraulic model parameters using synthetic aperture radar data: an integrated multidisciplinary approach. *IEEE Transactions on Geoscience and Remote Sensing* **47**: 455–467.

Peters NE, Freer J, Aulenbach BT. 2003. Hydrological dynamics of the Panola Mountain Research Watershed, Georgia. *Ground Water* **41**: 973–988.

Quinn P. 2004. Scale appropriate modelling: representing cause-and-effect relationships in nitrate pollution at the catchment scale for the purpose of catchment scale planning. *Journal of Hydrology* **291**: 197–217.

Rosen MR, Bright J, Carran P, Stewart MK, Reeves R. 1999. Estimating rainfall recharge and soil water residence times in Pukekohe, New Zealand, by combining geophysical, chemical, and isotopic methods. *Ground Water* **37**: 836–844.

Rupp DE, Selker JS. 2006. Information, artifacts, and noise in $dQ/dt - Q$ recession analysis. *Advances in Water Resources* **29**: 154–160.

Rupp DE, Woods RA. 2008. Increased flexibility in base flow modelling using a power law transmissivity profile. *Hydrological Processes* **22**: 2667–2671.

Savenije HHG. 2009. HESS opinions "The art of hydrology". *Hydrology and Earth System Sciences* **13**: 157–161.

Sidle RC, Noguchi S, Tsuboyama Y, Laursen K. 2001. A conceptual model of preferential flow systems in forested hillslopes: evidence of self-organization. *Hydrological Processes* **15**: 1675–1692.

Sidle RC, Tsuboyama Y, Noguchi S, Hosoda I, Fujieda M, Shimizu T. 2000. Stormflow generation in steep forested headwaters: a linked hydrogeomorphic paradigm. *Hydrological Processes* **14**: 369–385.

Sivapalan M, Grayson R, Woods R. 2004. Scale and scaling in hydrology. *Hydrological Processes* **18**: 1369–1371.

Sklash MG, Farvolden RN, Fritz P. 1976. A conceptual model of watershed response to rainfall, developed through the use of oxygen-18 as a natural tracer. *Canadian Journal of Earth Sciences* **13**: 271–283.

Soulsby C, Neal C, Laudon H, Burns DA, Merot P, Bonell M, Dunn SM, Tetzlaff D. 2008. Catchment data for process conceptualization: simply not enough?. *Hydrological Processes* **22**: 2057–2061.

Stewart MK, Fahey BD. 2010. Landuse effects on runoff generating processes in tussock grassland indicated by mean transit time estimation using tritium. *Hydrological Earth Systems and Sciences Discussions* **7**: 1073–1102.

Stewart MK, Mehlhorn J, Elliott S. 2007. Hydrometric and natural tracer (oxygen-18, silica, tritium and sulphur hexafluoride) evidence for a dominant groundwater contribution to Pukemanga Stream, New Zealand. *Hydrological Processes* **21**: 3340–3356.

Szilagyi J. 2009. Comment on "Power law catchment-scale recessions arising from heterogeneous linear small-scale dynamics" by C. J. Harman, M. Sivapalan, and P. Kumar. *Water Resources Research* **45**: W12601.

Tallaksen LM. 1995. A review of baseflow recession analysis. *Journal of Hydrology* **165**: 349–370.

Tetzlaff D, McDonnell JJ, Uhlenbrook S, McGuire KJ, Bogaart PW, Naef F, Baird AJ, Dunn SM, Soulsby C. 2008. Conceptualizing catchment processes: simply too complex? *Hydrological Processes* **22**: 1727–1730.

Thyer M, Renard B, Kavetski D, Kuczera G, Franks SW, Srikanthan S. 2009. Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: a case study using Bayesian total error analysis. *Water Resources Research* **45**: W00B14.

Tromp-van Meerveld HJ, McDonnell JJ. 2005. Comment to "Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes, *Journal of Hydrology* 286: 113–134". *Journal of Hydrology* **303**: 307–312.

Tromp-van Meerveld HJ, McDonnell JJ. 2006a. Threshold relations in subsurface stormflow: 1. A 147-storm analysis of the Panola hillslope. *Water Resources Research* **42**: W02411.

Tromp-van Meerveld HJ, McDonnell JJ. 2006b. Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis. *Water Resources Research* **42**: W02409.

Vache KB, McDonnell JJ. 2006. A process-based rejectionist framework for evaluating catchment runoff model structure. *Water Resources Research* **42**.

Wagener T, Wheater HS. 2006. Parameter estimation and regionalization for continuous rainfall–runoff models including uncertainty. *Journal of Hydrology* **320**: 132–154.

Western AW, Grayson RB, Blöschl G. 2002. Scaling of soil moisture: a hydrologic perspective. *Annual Review of Earth and Planetary Sciences* **30**: 149–180.

Western AW, Zhou SL, Grayson RB, McMahon TA, Blöschl G, Wilson DJ. 2004. Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes. *Journal of Hydrology* **286**: 113–134.

Western AW, Zhou SL, Grayson RB, McMahon TA, Blöschl G, Wilson DJ. 2005. Reply to comment by Tromp van Meerveld and McDonnell on spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes. *Journal of Hydrology* **303**: 313–315.

Wilson DJ, Western AW, Grayson RB, Berg AA, Lear MS, Rodell M, Famiglietti JS, Woods RA, McMahon TA. 2003. Spatial distribution of soil moisture over 6 and 30 cm depth, Mahurangi river catchment, New Zealand. *Journal of Hydrology* **276**: 254–274.

Wittenberg H. 1999. Baseflow recession and recharge as nonlinear storage processes. *Hydrological Processes* **13**: 715–726.

Woods RA, Grayson RB, Western AW, Duncan MJ, Wilson DJ, Young RI, Ibbitt RP, Henderson RD, McMahon TA. 2001. Experimental design and initial results from the Mahurangi River Variability Experiment: MARVEX. In *Observations and Modelling of Land Surface Hydrological Processes*, Lakshmi V, Albertson JD, Schaake J (eds). American Geophysical Union: Washington, DC; 201–213.

